Introduction to Monte Carlo Simulations

Scenarios, Breakthroughs, and Applications

Minwoo Yoo

Department of Economics
George Washington University

Econ 8376 - Spring 2026

The Goal: Knowing Your Estimator

We build an estimator $\hat{\theta}$ to learn about a parameter θ.
Before we trust $\hat{\theta}$, we need to know its properties:

- Is it **unbiased**? ($E[\hat{\theta}] = \theta$?)
- What is its **variance**?
- What is its **distribution**? (Normal? Chi-squared?)

Usually, we use **Asymptotic Theory** (Math) to answer these.

But what if the math fails?

Scenario A: The "Small Sample" Anxiety

The Situation:

- You derived that $\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N(0, V)$ as $n \to \infty$.
- But in your actual research, you only have **30 observations** ($n = 30$).

The Obstacle:

- Is $n = 30$ large enough? Is the approximation accurate?
- Or is the finite-sample distribution actually skewed?

Scenario A: The "Small Sample" Anxiety

The Situation:

- You derived that $\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N(0, V)$ as $n \to \infty$.
- But in your actual research, you only have **30 observations** ($n = 30$).

The Obstacle:

- Is $n = 30$ large enough? Is the approximation accurate?
- Or is the finite-sample distribution actually skewed?

The Breakthrough

Don't guess. Simulate.

1. Create 10,000 artificial worlds with $n = 30$.
2. Calculate $\hat{\theta}$ in each world.
3. **Plot the histogram.** If it's not Normal, the asymptotic theory is misleading you.

Scenario B: The "Impossible Integral"

The Situation (Structural Models / Bayesian):

- You need to calculate an expected value (e.g., Market Shares in BLP).
- This involves a high-dimensional integral:

$$\sigma_j(p) = \int \cdots \int P_j(p, \nu) f(\nu) d\nu$$

The Obstacle:

- There is **no closed-form solution**. You cannot solve it by hand.

Scenario B: The "Impossible Integral"

The Situation (Structural Models / Bayesian):

- You need to calculate an expected value (e.g., Market Shares in BLP).
- This involves a high-dimensional integral:

$$\sigma_j(p) = \int \cdots \int P_j(p, \nu) f(\nu) d\nu$$

The Obstacle:

- There is **no closed-form solution**. You cannot solve it by hand.

The Breakthrough (Monte Carlo Integration)

Transform Calculus into Arithmetic. By the Law of Large Numbers (LLN), sample average converges to expectation:

$$\int g(\nu) f(\nu) d\nu \approx \frac{1}{R} \sum_{r=1}^{R} g(\nu_r)$$

Just draw random ν_r and take the average.

Scenario C: The "Black Box" Estimator

The Situation:

- Your estimator $\hat{\theta}$ is the result of a complex algorithm or optimization (e.g., a complicated GMM or ML estimator).
- You cannot easily derive the formula for its Variance.

The Obstacle:

- How do I report the Standard Error?

Scenario C: The "Black Box" Estimator

The Situation:

- Your estimator $\hat{\theta}$ is the result of a complex algorithm or optimization (e.g., a complicated GMM or ML estimator).
- You cannot easily derive the formula for its Variance.

The Obstacle:

- How do I report the Standard Error?

The Breakthrough

Empirical Variance.

- Run the algorithm 1,000 times on simulated data.
- Calculate the variance of the 1,000 results.
- That is your approximation of the true variance.

So, What is This?

This approach of "solving problems by generating random numbers" is called:

Monte Carlo Simulation

It replaces **analytical derivation** (which might be hard or impossible) with **computational repetition** (which is easy for computers).

The Experiment Design: Testing the CLT

The Question: Standard theory (CLT) says $\bar{Y}$ should be Approx. Normal even if N is finite. Is $N = 25$ "large enough" for this to hold?

The Blueprint: To verify this, we will act as the **Data Generating Process (DGP)**.

1. **Assumption:** The true population is Standard Normal ($Y \sim N(0, 1)$).
2. **Action:** Draw $N = 25$ random observations from this population.
3. **Calculation:** Compute the sample mean $\bar{Y}$.
4. **Repetition:** Do this 100,000 times to approximate the distribution of $\bar{Y}$.
5. **Verification:** Check if the resulting histogram looks like a Bell Curve.

Example: Setup the Estimator

Let's translate our blueprint into Python.

```python
import numpy as np
import matplotlib.pyplot as plt

# 1. Define the "Truth" (DGP Parameters)
n = 25        # Small sample size
muY = 0       # True Mean
sigY = 1      # True Std Dev

# 2. Define the Estimator (The "Rule")
def get_sample_mean(n, muY, sigY):
    # Generate data from the true distribution
    Y = sigY * np.random.randn(n) + muY

    # Return the estimate (Sample Mean)
    return Y.mean()
```

Example: The Simulation Loop

```
# 3. The Monte Carlo Loop (Repeat 100,000 times)
MC = 100000
sample_means = np.zeros(MC) # Container for results

for i in range(MC):
    # Store the result of each experiment
    sample_means[i] = get_sample_mean(n, muY, sigY)

# 4. Analysis
print("Exp. Value:", sample_means.mean()) # Check Unbiasedness
print("Variance:", sample_means.var())   # Check Efficiency
```

Example: Visualization Code

Finally, we draw the histogram to see the distribution.

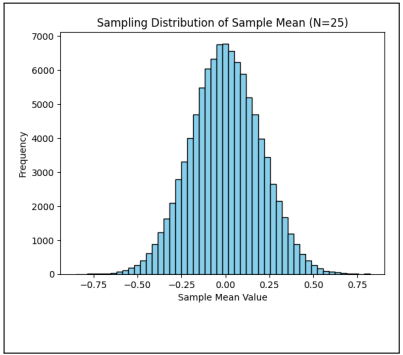
```
# 5. Plot the sampling distribution
plt.hist(sample_means, bins=50, color='skyblue', edgecolor='black')

plt.title("Sampling Distribution of Sample Mean (N=25)")
plt.xlabel("Sample Mean Value")
plt.ylabel("Frequency")

# Save or Show the plot
plt.show()
```

The Result: Visual Proof

- We generated 100,000 **Sample Means**.
- The histogram shows the **Sampling Distribution**.
- **Conclusion:** Even with $N = 25$, the distribution of $\bar{Y}$ looks fairly Normal (Bell-shaped).
- The CLT works reasonably well here!



Comparison: Monte Carlo vs. Bootstrap

	Monte Carlo	**Bootstrap**
When to use?	To study properties of an estimator (e.g., Does $\bar{Y}$ work well?).	To calculate SEs/CIs for *your specific dataset*.
The "World"	We create the world (DGP is known).	The sample is the world (DGP is unknown).

Summary & Best Practices:

- Use MC when math is too hard or N is too small.
- Always set a seed ('np.random.seed') for reproducibility.
- Ensure enough repetitions ($R \geq 1000$).